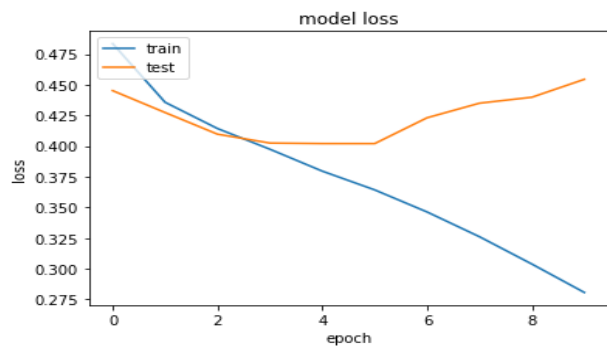
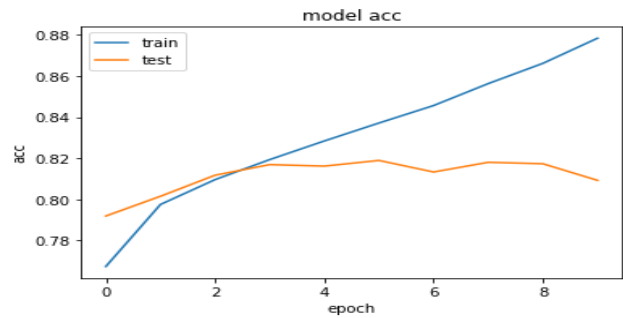
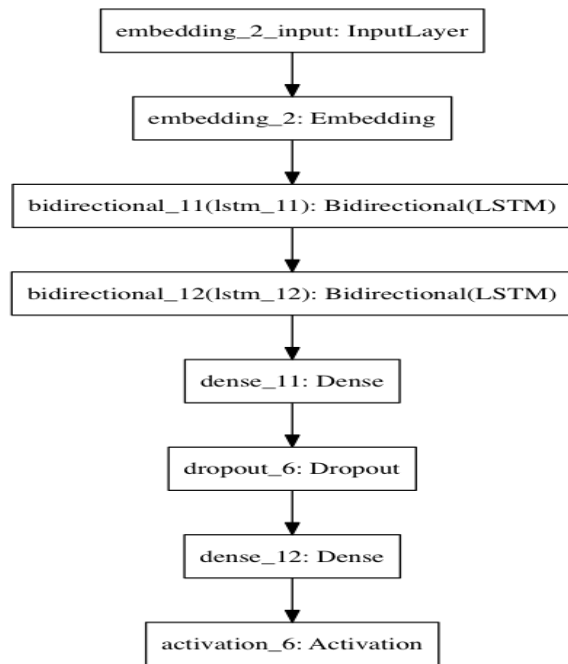


學號：R06942098 系級：電信碩一 姓名：曾柏偉

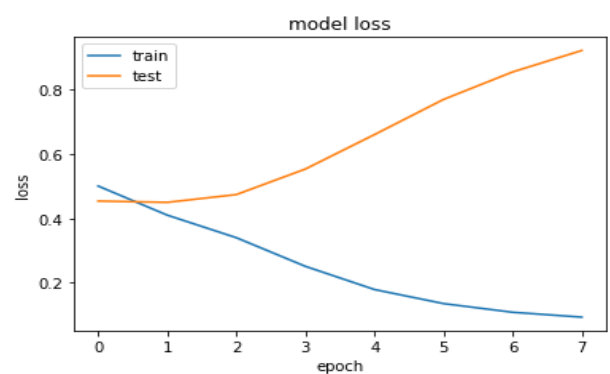
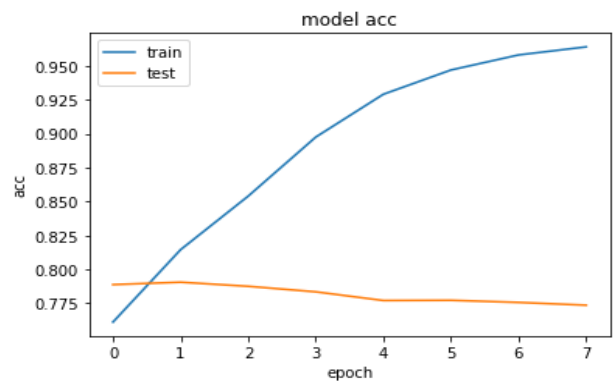
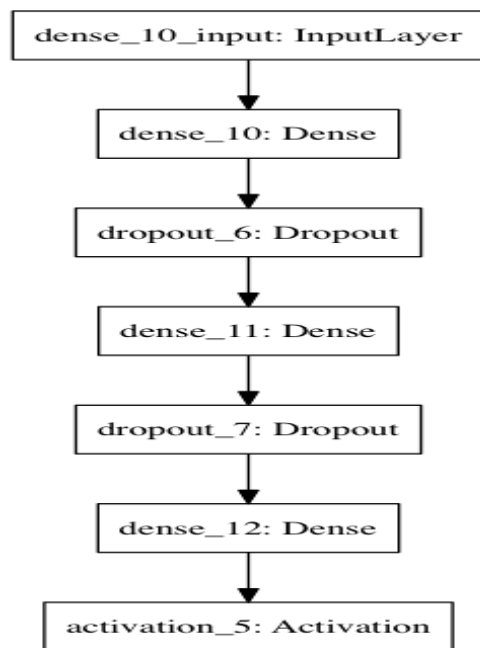
1. (1%) 請說明你實作的 RNN model，其模型架構、訓練過程和準確率為何？
(Collaborators:)

答：我有使用 gensim 的 word2vec，並使用原本的 train data 組成一個 embedding matrix 並且 fix weight 去 training，得到了 0.82006 是最好的 accuracy。



2. (1%) 請說明你實作的 BOW model，其模型架構、訓練過程和準確率為何？
(Collaborators:)

答：Bag of word 實作上其實很快，只接了二層的 fs，以下是 performance



3. (1%) 請比較 bag of word 與 RNN 兩種不同 model 對於"today is a good day, but it is hot"與"today is hot, but it is a good day"這兩句的情緒分數，並討論造成差異的原因。

(Collaborators:)

答：bag of word：對於 bag of word 都得到 0.75312412

RNN：‘today is a good day , buy it is hot’得到 0.97791505

‘today is hot , but it is a good day’ 得到了 0.97807354

model	Today is a good day, but it is hot	Today is hot , but it is a good day.
RNN	0.97791505	0.97807354
Bag of word	0.75312412	0.75312412

4. (1%) 請比較"有無"包含標點符號兩種不同 tokenize 的方式，並討論兩者對準確率的影響。

(Collaborators:)

答：有 filter 掉標點符號的 accuracy : 0.82006

無 filter 標點符號的 accuracy : 0.80632 ，兩者的這個差距在這次作業算是很大的。

Term	accuracy
有標點符號	0.82006
無標點符號	0.80632

5. (1%) 請描述在你的 semi-supervised 方法是如何標記 label，並比較有無 semi-supervised training 對準確率的影響。

(Collaborators:)

答：在 self-training 的過程中，我使用 softmax 把機率分成兩個數值，每一次 predict 時我只取各機率大於 0.95 的 unlabeled data，並 concatenate 進去原本的 training data 內，我反覆執行直到全部 unlabeled data 都被 sample 完畢，然後準確率提到 0.82556 了！！